Project Gutenberg: Free "Books" Online

BY KEVIN M. SAVETZ

What if books were free? What if anyone could get a copy of whatever books he or she wanted? Would libraries become extinct? Would bookstores crumble?

When Johann Gutenberg invented movable type in the 1400s, the cost of books plummeted to a fraction of their former prices, and thousands of hours of production time were saved. For the first time, the general public had access to books.

Michael Hart wants to lead a similar revolution.

ELECTRONIC GUTENBERG

In 1971, Hart was given $100,000,000 in computer time by the operators of the Xerox Sigma V mainframe at the University of Illinois. In an effort to "earn" his free computer time, Hart typed in the Declaration of Independence and attempted to distribute it across the network so everyone could have one.

With that act, Project Gutenberg was born. Today, this nonprofit organization releases electronic versions of public-domain literature. Its goal is to give away one hundred billion electronic texts by the end of 2001—10,000 titles, each distributed to 100 million people. "There are predicted to be a billion people on the networks by 2001. If we hit 10% of them, that's a trillion," says Hart, executive director of Project Gutenberg.

The Project Gutenberg volunteers release an electronic edition of a work a short time after that book reaches the public domain. Each document is in plain-vanilla ASCII, so that everyone using anything from Apples to Z-80s is able to read them. Fancy graphics and text formatting are spurned in favor of small, easily shared documents.

Titles already released include classic literature such as *Aesop's Fables, Alice in Wonderland, Moby Dick,* and works of Shakespeare; historical documents such as the Magna Carta, US Declaration of Independence, and the US Constitution; reference books such as *Roget's Thesaurus,* almanacs, census data, the *CIA World Fact Book,* and mathematical constants; and religious documents such as the Bible and the Book of Mormon.

Why would anyone want to read a book on a computer rather than... well, from a book? "The biggest advantage of electronic books," Hart says, "is they cost about one one-hundredth of what a book costs. If you buy a copy of *Alice in Wonderland* right now, even in paperback, it will cost at least five bucks. On a 1.44-megabyte floppy, the same book uses a nickel's worth of space."

"The biggest advantage of electronic books," Hart says, "is they cost about one one-hundredth of what a book costs. If you buy a copy of *Alice in Wonderland* right now, even in paperback, it will cost at least five bucks. On a 1.44-megabyte floppy, the same book uses a nickel's worth of space."

Because electronic texts (or "e-texts") are stored on a computer, it is much simpler to search a book—or an entire library—for specific information. Research using e-texts is faster and easier than using traditional media. "My favorite example," Hart says, "is a paper I wrote on death and marriage in *Hamlet, Macbeth, Romeo and Juliet,* and *Othello.* This is the kind of paper you couldn't even write without a computer. There are too many citations for even a Shakespeare scholar—you would be up to the ceiling in index cards.

"I am hoping—presuming almost—that as books are added to the public domain electronic library, these kinds of papers will be written. Your everyday research paper is 90% research and 10% writing. With e-texts, it will be 10% research and 90% writing the paper," he says. Research using e-texts doesn't require trips to the library and allows more time for thinking about what to write rather than shuffling pages. According to Hart, because the actual research will take less time, researchers will have more time for original thought.

"The term 'exhaustive research' will disappear, because it won't exhaust you any more," he says. "Painstaking research will be as far in the past as painstaking hand copying in the monasteries."

Although the idea of reading books on a computer screen may seem farfetched now, Hart believes that as the cost of storing information and accessing networks continues to fall, electronic texts will become a standard method of reading and retrieving information. "I'm not saying books are going to completely die out in the next ten years, but this is an idea whose time has come," Hart says.

So far, the volunteer staff has released about seventy texts in electronic form. Hart says the number of volunteers who type or scan books triples every year. There are currently about two hundred volunteers. Hart

Copyright © 1993 by Kevin M. Savetz. All rights reserved. *Kevin Savetz is a free-lance writer based in Arcata, California. His email address is "savetz@rahul.net". His two cats are of the Cheshire variety.*

has only met about 5% of them—the rest are network users near and far. For the past three years, the project has released twice the number of books as the previous year. "This doubling curve will get us to our goal of 10,000 books by 2001," Hart says.

Obviously, libraries aren't extinct yet. Although they are a boon for research and searching, electronic texts do have drawbacks for the casual reader. You can't read one in the bath, and you might hesitate to tote a portable computer with you on public transit, just to catch the last chapter of *Moby Dick*. Still, Hart is betting on the near future, when storage will be cheap, computers will be even more portable, and screens will be easier to read. By the time the 10,000th e-text rolls off the presses in the year 2001, this will undoubtedly be the case.

One technical problem slowing the release of texts lies in copyright law. Unless given specific permission, Project Gutenberg may only reproduce works that are in the public domain—uncopyrighted. Under current copyright law, works don't enter the public domain until the author has been dead for 50 years, so don't expect too many best sellers to be given away on the Internet archives.

Although the idea of giving free books to the masses is an honorable one, will Project Gutenberg's efforts have a similar effect to Johann Gutenberg's? Johann's efforts made inexpensive books available to the general public. Project Gutenberg has made books even less expensive—nearly free, even considering transfer time and storage space. However, for the time being, the general public cannot collect on Project Gutenberg's promise because those of us with access to home computers and online services are in the minority. Until everyone uses the Internet or has a CD-ROM drive, Project Gutenberg cannot have the impact that Johann Gutenberg's movable type did over five hundred years ago.

ACCESSING THE PROJECT GUTENBERG ARCHIVES

For further information, or to retrieve electronic books via FTP (the Internet's file transfer protocol), anonymous FTP directly to the Project Gutenberg archives: *mrcnext.cso.uiuc.edu (128.174.201.12)*. Texts are available in directories sorted by the year of release: *etext/extext91, etext/etext92,* and *etext/etext93*. For more information about Project Gutenberg and the current newsletter, look in directory *etext/articles*.

Project Gutenberg files are available from other FTP sites as well: *nptn.org (192.55.234.52)* in */pub/e.texts/gutenberg ftp oes.orst.edu (128.193.124.2);* in */pub/almanac/etext quake.think.com (192.31.181.1);* in */pub/etext91* or */pub/etext92 simtel20.army.mil* or *(192.88.110.20);* in *pd1:<msdos.books> wuarchive.wustl.edu (128.252.135.4)*; and in */mirrors/msdos/books*.

The Almanac Information Server, located at Oregon State University, allows text retrieval via electronic mail. To retrieve a file via email, first send the following line by itself to *almanac@oes.orst.edu:*

send gutenberg catalog

This will get you instructions on how to send further requests, and will list the available files. For example, to retrieve *Alice's Adventures in Wonderland,* send to *almanac@oes.orst.edu:*

send gutenberg alice

Email questions about Project Gutenberg to: *dircompg@ux1.cso.uiuc.edu*.

E-texts released in 1993, in directory */etext/etext93* (check archive site for current filenames—xxx's denote version numbers which change periodically):

A Princess of Mars [Mars #1](pmars10x.xxx)
Aladdin and the Magic Lamp (alad10xx.xxx)
Anne of the Island (iland10x.xxx)
Black Experience (Coombs) (blexp10x.xxx)
Civil Disobedience, Thoreau (civil10x.xxx)
Clinton's Inaugural Address (clintonx.xxx)
Descartes' Reason Discourse (dcart10x.xxx)
Email 101 by John Goodwin (email025.xxx)
Gods of Mars [Mars #2] (gmars10x.xxx)
LOC Workshop on Etexts (locet10x.xxx)
Milton's Paradise Regained (rgain10x.xxx)
NREN, by Jean Armour Polley (nren210x.xxx)
Pi (circumference/diameter) (pimil10x.xxx)
Red Badge of Courage, Crane (badge10x.xxx)
Surfing the Internet (Surf10xx.xxx)
The 32nd Mersenne Prime (32pri10x.xxx)
The Communist Manifesto (manif10x.xxx)
The Dawn of Amateur Radio (radio10x.xxx)
The First 100,000 Primes (prime10x.xxx)
The Marvelous Land of Oz (ozland10.xxx)
The Number "e" [Natural Log](ee610xxx.xxx)
The Scarlet Pimpernel (scarp10x.xxx)
The Square Root of Two (2sqrt10x.xxx)
The Wonderful Wizard of Oz (wizoz10x.xxx)
The World Factbook (world192.xxx)
Thuvia, Maid of Mars [#4] (mmars10x.xxx)
Tom Sawyer [Wiretap/Twain] (sawyr10x.xxx)
Warlord of Mars [Mars #3] (wmars10x.xxx)
What Is Man? Mark Twain (wman10xx.xxx)

E-texts released in 1992, in directory /etext/etext92

1991 CIA World Factbook (world91a.xxx)
A Christmas Carol (Dickens) (carol10x.xxx)
Aesop's Fables (Advantage) (aesopa10.xxx)
Anne of Avonlea (avon10xx.xxx)
Anne of Green Gables (anne10xx.xxx)
Data From the 1990 Census (uscen901.xxx)
Dr. Jekyll and Mr. Hyde #1 (hyde10xx.xxx)
Dr. Jekyll and Mr. Hyde #2 (hydea10x.xxx)
Far From the Madding Crowd (crowd13x.xxx)
Frederick Douglass (duglas10.xxx)
Herland (for Mother's Day) (hrlnd10x.xxx)
Hitchhiker Internet Guide (hhgi10xx.xxx)
New Etext of Bible (KJV) (bible10x.xxx)
NorthWestNet NUSIRG Guide (nusirg10.xxx)
O Pioneers! Willa Cather (opion10x.xxx)
Paradise Lost (Raben) (plrabn11.xxx)
Song of the Lark (Cather) (song10xx.xxx)
Sophocles' Oedipus Trilogy (oedip10x.xxx)
The 1990 US Census (2nd) (uscen902.xxx)
The Gift of the Magi-O Henry (magi10.txt)
The Hackers' Dictionary (jargn10x.xxx)
The Legend of Sleepy Hollow (sleep10x.xxx)
The Scarlet Letter (scrlt10x.xxx)
The Time Machine-HG Wells (timem10x.xxx)
The War of the Worlds-HGW (warw10xx.xxx)
Zen & the Art of Internet (zen10xxx.xxx)

E-texts released in 1991, in directory /etext/etext91

1990 CIA World Factbook (world12x.xxx)
Aesop's Fables (aesop11x.xxx)
Alice in Wonderland (alice29x.xxx)
Moby Dick (mobyxxxx.xxx)
Paradise Lost (plboss11.xxx)
Peter Pan (for US only) (peter14a.xxx)
Roget's Thesaurus (roget12x.xxx)
The Book of Mormon (mormon13.xxx)
The Federalist Papers (feder12x.xxx)
The Hunting of the Snark (snark12x.xxx)
The Song of Hiawatha (hisong11.xxx)
Through the Looking Glass (lglass16.xxx)

BOOKS ON CD-ROM

In July, Walnut Creek CD-ROM released a CD-ROM containing all of Project Gutenberg's current collection of e-texts. The CD-ROM conforms to ISO-9660, which is the most common and generic format. It will work with almost any system. All the files are in plain ASCII, so that they can be used with the widest variety of computers and software.

The price is $39.95 (plus sales tax in California) plus $5 shipping and handling per order. For more information, contact Bob Bruce, Walnut Creek CD-ROM, 1547 Palos Verdes, Suite 260, Walnut Creek, CA, 94596. 1-800-786-9907. Or send electronic mail to "orders@cdrom.com".

The company plans to update the CD-ROM with new e-texts every six months. Updates will be available by subscription. Part of the proceeds from the sale of the disc go to Project Gutenberg to help pay for the creation of new e-texts. The company offers a variety of CD-ROMS—for a list, you can FTP the latest catalog from ftp.cdrom.com:/cdrom/catalog, or send email to "info@cdrom.com". ■